Darwin s invisible hand

Bosdriesz, E.

2015

CHAPTER 7

General discussion

A little over four years ago, I started as a graduate student in systems biology. As I had a background in physics rather than in biology, my knowledge of how cells function was limited, and I had to learn a lot of biochemistry, cell biology, microbiology, molecular biology, and much more. Acquainting myself with these topics was equally exciting and intimidating. Here we have a system, consisting of many thousands of different kinds of interacting macromolecules. Some of these molecules are well characterised, other not so well, but all of them are complex molecular machines themselves. Together, they somehow give rise to a functioning cell. The sheer complexity and ingenuity of such a system of macromolecules seemed as beautiful as it seemed incomprehensible.

I mention my background and my first experiences in the field of systems biology because they have shaped my view on cellular biology and the way to study it. In this discussion I will try to elaborate on this view, and how I believed it influenced the approach I've taken. My views on this are not necessarily new or original. In fact, many of the issues I will address in this discussion were already addressed many years ago by people like Max Delbrück [210] and Gunther Stent [211]. Moreover, these views are definitely not the only "correct" views. However, stating them explicitly will make the overarching themes of this thesis more clear, I hope.

It is not straightforward to define a single, central question that is addressed in this thesis. In fact, I address several problems that may at first sight appear quite unrelated: general properties of metabolic networks, the kinetics of nutrient uptake systems, and the regulation of ribosome expression in *Escherichia coli*. While all these questions pertain to cellular functioning, they are not evidently functionally related to each other. Rather, the relation between the different problems addressed here is to be found in the approach and methods used, which are influenced by my training in

physics.

The history of physics is characterised by the quest for unification and generalisation. Seemingly different phenomena turn out to be closely related at a deeper level, or conversely, descriptions of certain phenomena can be used to describe other, (apparently) unrelated phenomena. Prime example are the laws of Newton that describe both celestial and terrestrial mechanics, the laws of Maxwell that showed that magnetism and electricity are principally the same, statistical physics which describes thermodynamics in term of mechanics, and so forth. As a result, a physicist is educated with the idea that the more we know about some phenomena, the "simpler" its description should become. For instance, in the 1950s and 60s, with the development of increasingly powerful particle accelerators, ever more subatomic particles were identified. This led to the so called "particle zoo", a confusingly long list of "elementary" particles. However, all these particles turned out to consist of a limited number of more fundamental building, called quarks, the discovery of which greatly simplified the subatomic world*.

Typically, the most successful theories in physics take the form of general laws or principles. The situation in biology appears quite different. It is often thought that due to the dynamics of evolution every biological phenomenon is a historical "accident", and that this prohibits discovering universal "laws" in biology. While it is true that in biology we will probably never find a law as fundamental and quantitative as, say, the theory of general relativity, biology can in my opinion greatly benefit from the search for general principles. Most of the work in this thesis is aimed at uncovering such general principles in the functioning of cells.

7.1 The search for general principles of cells

General principles of biomolecules

Genetics, molecular biology and biochemistry have been extremely successful in the identification and characterisation of the basic principles of the macromolecules found in living cells. Famous examples are: the double helix structure of DNA, the central dogma, DNA replication, translation, transcription, allostery, cooperativity, classification of enzyme mechanisms, the operon structure of prokaryotic genes, protein domain structure, chromatin organisation, sigma-factor mediated transcriptional control, two-component signal transduction in bacteria, and many, many more. It is often not emphasised enough in these molecular disciplines that the search for these principles is what organised those fields, brought their successes, and led to their major contributions to biology. Clearly, these principles are not restricted to any particular organism and there is a remarkable unity in the basic chemistry of life. So, in contrast to what is often said, the molecular disciplines have, in terms of their approach, quite a lot in common with physics.

*Quarks were originally introduced (independently) by Murray Gell-Mann and Georg Zweig only as a purely theoretical concept aimed to bring order in the particle zoo.

Why the same molecular principles are found to hold across species, regardless of their evolutionary relatedness, addresses deep evolutionary questions. They will most likely have to do with the importance of horizontal gene transfer in bacteria, the prevalence of parallel and convergent evolution, and the fact that all life eventually descends from a common ancestor. But I do not want to tell an evolutionary story here, it is the identification of these principles that is the topic of this thesis.

From general principles of molecules to general principles of molecular networks

About a hundred years ago when molecular and structural biology was initiated many researchers thought that the identification of the structures of all biological macromolecules would lead to an "understanding" of life.[†] However, life is all about the systems properties generated through molecular interactions: in order to generate energy and to grow, a cell needs to catalyse a vast number of biochemical conversions, organised in a metabolic network. In order to adapt to changing environmental conditions a cell needs to sense and process information about the environment, which is mediated by signalling networks. This information should lead to the desired response, which is executed by gene regulatory networks, and so on.

Searching for principles underlying the organisation of these molecular networks is a natural continuation of the successes reached with searching for principles of biological macromolecules. This is not a new insight but only started to have an impact on mainstream biology as soon as the measurement technologies were introduced that could inform us about the dynamics and wiring of these networks. This led to the field of systems biology, which aims at understanding the principles of molecular networks - in the same vein as the more "molecular" disciplines focussed on molecular principles. However, the principles at the network level turned out to be very different from those at the molecular level. They arise not solely from the properties of the molecules that constitute the network, but from the interactions between them. Moreover, even the function of these networks, such as robustness or sensitivity to perturbations, adaptability, response time, and homeostasis, are often properties of the network as a whole, and cannot be attributed to any particular single molecule or output parameter. Finally, we do not know all the kinetic properties of all the molecules making up the cell, and most likely will not know soon or ever. Our descriptions of networks are therefore inevitably incomplete, which appears to be a fundamental problem in systems biology. So, how then should we go about elucidating the principles of biological networks?

In a highly influential paper titled "More is Different" [212], Anderson addressed this problem more generally. He argued that "the reductionist hypothesis does not by any means imply the 'constructionist' one: The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct

[†]There has also always been a large school of molecular biologists who believed that an understanding life cannot come from solely understanding the physiochemical properties of biomolecules. See they essay by Stent for an interesting discussion on the (early) history of molecular biology [211].

the universe. [...] The constructionist hypothesis breaks down when confronted with the twin difficulties of scale and complexity." In his view, the sciences can be roughly organised in a linear hierarchy of increasing complexity, where the science of each additional layer obeys the laws of the underlying ones. Many body physics needs to be consistent with the laws form particle physics; in turn, chemistry needs to be consistent with the laws of many body physics, and so on. However, when going up a level of complexity or scale, new laws and concepts will need to be formulated that are different, but not less fundamental, as the laws governing the underlying science. A famous example is the physics of phase transitions, which is ultimately derived from the interactions of the particles but which nonetheless requires a fundamentally different description. Translated to cellular or microbial biology, it is not possible to simply "derive" the behaviour of a cell from the properties of the molecules that it is made of [‡], but it will require new concepts and generalisations.

What we learned from studies of macromolecules is that principles of macromolecules, such as protein allostery or cooperativity, do not depend on all the amino acids making up the protein or that promoter design of genes is independent of the DNA sequence of the coding genes. Likewise, not all kinetic properties of molecules matter for understanding networks. This realisation partially solves the problem of not knowing all kinetic parameters, and it suggests that a strategy of searching for the key molecular interactions and parameters of a network - akin to the order parameters in statistical physics of phase transitions - is a fruitful thing to do. The origin of this behaviour lies in the nonlinear dynamics of molecular networks. Nonlinear systems have the property that the importance of parameters changes with the state of the system, and that for particular behaviours only a subset of parameters is relevant. Therefore, reduced system descriptions - with mathematical models - in terms of this small set of essential parameters gives rise to a dynamic mechanism, e.g., for membrane transport, that has properties found across biological species.

This realisation leads to the concept of design principle (for lack of a better word): a particular set of molecular properties required for a certain functional network behaviour. So, for oscillations one often needs negative feedback, positive feedback leads to bistability and signalling cascades can give rise to sensitivity amplification. All cells are made from the same components, e.g., nucleotides, lipids, and amino acids organised into macromolecules. The intracellular milieu of cells is thus quite similar and the same physical constraints occur on biochemical processes inside cells of different species. In addition, cells have to carry out similar tasks: sensing, signal integration, induction of a response, metabolism, macromolecular synthesis, regulation of metabolic pathways, etcetera. Taken together, these two commonalities across species imply that they will often evolve the same molecular mechanisms when confronted with the same selective pressures. I am aware that this reasoning is very qualitative, and that undoubtedly examples can be found where this is not true, but I do believe this to be a strong organisation principle in biology.

[‡]Interestingly, Anderson also challenged the arrogance of molecular biologists who reduce life (specifically, the human organism), to "only" chemistry.

7.2 Aim and approach taken in this thesis

When thinking about design principles, the first question to answer is what the function of a system might be. Without an answer to this question, it is not possible to reason about the underlying logic of the system. In this thesis, I started from the proposition that (unicellular) organisms are subjected to strong selection for growth rate, a common assumption in microbiology. As a consequence they need to efficiently allocate limited resources, so as to minimise doubling time, and it is this resource allocation problem that shapes the cellular physiology. In my thesis, I have used these premises to discover design principles, and applied it to several different problems:

1. General characteristics of optimal metabolic states, both in a batch environment and in a chemostat setting. This was the topic of chapters 2 and 3.
2. Low affinity and binding protein dependent membrane transporters, which was the topic of chapters 4 and 5, and finally,
3. *Escherichia coli*'s mechanism for growth rate tuning to its environment.

In the remainder of this discussion, I will take a bird's eye view on my own work, discuss its broader implications and address the advantages and limitations of the "design principle" approach.

General characteristics of optimal metabolic states

In chapters 2 and 3 the aim was to define general characteristics of metabolic states that optimise the specific growth rate, which was defined as the specific biomass synthesis flux. I believe that the relevance of these results will mainly be as "tools" to reason about large (kinetic) models where intuition alone does not suffice. An example of this is the result that in the chemostat model the eventual outcome of prolonged evolution must be a metabolic strategy that is an elementary flux mode. However, coexistence of two strains that employ a different metabolic strategy can occur, as well as the speciation of a species initially employing a mixed metabolic strategy into two "specialists". In other words, overflow metabolism, the simultaneous use of a respiratory and fermentative pathway, cannot be optimal in a chemostat (model). However, a coexistence of a respiring and a fermenting strain can be an evolutionary stable state. Once established, this seems quite intuitive, but for me it only became apparent after we established the general characteristics of optimal states.

The results obtained in these chapters are mathematical derivations, so, provided that the underlying assumptions are valid, they are strictly true. For instance, *given the underlying assumptions* about cellular growth rate made in chapter 3, an optimal metabolic strategy *has to be* an elementary flux mode. The problem is, of course, that the underlying assumptions are a sensible approximation of cellular growth at best (and nonsense in the worst case). Perhaps the most important assumption is that what we define as the specific growth rate, the rate of protein formation per unit biomass, is the subject of evolutionary optimization in microorganisms. What

is crucial in our analysis is the assumption the increase of the concentration of a certain enzyme decreases the concentration of another enzyme (or increases the size of the cell). This assumption is based on the observation that the protein density of cells is quite constant across conditions [151]. Related to this, the assumption implies that the process of translation has considerable control over the growth rate, which is supported by the experimental finding that the cost of protein synthesis is in translation [20]. This is important because it gives rise to the resource allocation problem. If the concentration of an enzyme can be increased "for free", there is no optimal enzyme concentration since a catalytic benefit can be gained at no cost.

What we have not taken into account, is that the membrane surface-to-volume ratio might be limiting. Under particular conditions a cell might become nearly fully "uptake limited", which means that (biosynthetic) fluxes are not limited by the capacity of metabolic pathways, but solely by the capacity to take up nutrients. In other words, most or all growth-rate control will reside in the transport reactions. Under these conditions, the problem will be allocation of limited membrane area resources, rather than biosynthetic capacity. This will have considerable consequences for characteristics of optimal states, also of those parts not directly related to nutrient uptake. For instance, when due limitations of membrane area most, but not all, flux control resides in the transport steps, it will be advantageous for the cell to keep the "product" of the transporter, the intracellular substrate, very low in order to minimise product inhibition of and substrate efflux by the transporter. This can be achieved by investing a lot in the first steps of the pathways metabolising this substrate, which in turn will require redirection of resources from other parts of the metabolic network. Hence, a natural and important extension to the theory of optimal metabolic states will be to incorporate membrane area limitations.

Principles of membrane transporters

Chapters 4 and 5 dealt with design principles of membrane transporters. The premise was that membrane bound proteins not only occupy biosynthetic resources, but also membrane area resources, and that therefore the pressure to use them efficiently is especially strong. Whether or not membrane area is really a limiting resource is still an open question. There is evidence that suggests that the membranes of bacteria are "full": estimations indicate that in *E. coli* a significant fraction of membrane area is required for glucose uptake alone [100], over-expression of membrane proteins is toxic to cells [103] and when nutrients are scarce and cells grow slower, they reduce in size and hence increase their surface-to-volume ratio[§] [101, 102]. There are also theoretical reasons to expect that cells minimise their surface area [15][¶]. However, direct evidence showing that the membrane area is a limited resource is still lacking.

The aim of chapter 4 was to "explain" the logic behind the existence of low affinity membrane transporters: low affinity transporters can enhance the net uptake rate

[§]The extent to which a reduction in cell size increases the surface-to-volume ratio depends on cell shape.

[¶]These have to do with the fact that the membrane itself does not contribute to growth but it does require metabolic resources to synthesise it. Therefore, "investments" in it should be minimised.

by reducing substrate efflux. This was specifically studied for transport by means of facilitated diffusion, a common transport mechanism. However, substrate efflux will only pose a problem to a cell for particular intracellular substrate concentrations, and these depend on the interplay between the kinetics of uptake and metabolism. In the case of glucose uptake in yeast, a simulation of a detailed kinetic model of transport and glycolysis indicated that the intracellular glucose concentration is expected to be high enough for this effect to be relevant, but an experimental test of this hypothesis obviously remains necessary. It is unclear to what extent this hypothesis can be extended to active transport uptake systems such as symporters or antiporters.

In chapter 5, the role of substrate binding proteins in transport was studied. Our results suggest that binding proteins can enhance frequency of interactions between transporters and substrates, which is partially achieved by an increase in the effective substrate concentration. This was shown for a very general model. It is known that some binding proteins are also involved in processes such as nutrient sensing, and it is thus likely that binding proteins can also enhance the frequency of interactions between receptors and ligands. On the other hand, the time it takes for ligands to bind to, and, more importantly, be released by, the binding protein will also surely affect signalling dynamics. It will be interesting to see what the implications of binding proteins are on chemotaxis and signalling. This would require extending our analysis beyond the steady state situation, to see the effect of binding proteins on the dynamics of ligand-receptor interaction upon a change in ligand concentration.

Principles of growth rate tuning by the ribosome concentration across fast-growing microorganisms

In chapter 6, about the regulation of ribosome expression in *E. coli*, a simplified model allowed for identification of the critical features of the stringent response system, that enable *E. coli* to attain maximal growth rates in a very robust manner. This simplified model can function as a template for studying regulation of ribosome expression in other organisms, also when the molecular implementation is quite different. We proposed that the regulation of ribosome expression in *Bacillus subtilis* and *S. cerevisiae* might be such examples. It would be interesting to study these regulatory mechanisms in more detail, and also look systematically at the regulation of ribosome expression in other organisms. More generally, the proposed feedback scheme can be a strategy for the optimal resource allocation of any enzymatic system where multiple pathways converge and the intermediates they produce are consumed by a common process.

The ability to optimally and robustly regulate ribosome expression ultimately depends on a separation in scales of the affinity of the ribosomes for amino-acids and their product-inhibition strengths. This gives rise to ultrasensitive behaviour of uncharged tRNA bound to the ribosomal A-site upon small deviations from their optimal ribosome concentrations. This ultrasensitivity is most likely difficult to directly test experimentally because it would require precise manipulation of the ribosome concentration in an ppGpp independent manner. However, it should

be quite straightforward to test the resulting consequences. That the wild-type ribosome concentration is optimal for growth has been shown for one particular growth condition in *E. coli* [26]. It would be interesting to test this for other growth conditions and other fast-growing organism as well. The robustness of the system could be tested by manipulating e.g., the RelA concentration in the cell. According to our model, moderate changes should have no effect on the ribosome concentration and the attained growth rate. While this would not be a direct proof that these conclusions derive from an underlying ultrasensitivity around the optimum, it would strongly suggest so.

7.3 Concluding remarks: Advantages and limitations of taking design principles as a guide

As design principles are by definition abstractions of more complex systems, it is often not straightforward to experimentally test them. For instance, a finding of the form *"Metabolite X allosterically regulates enzyme B"* is typically easier to validate or falsify than a finding of the form *"Proteins A, B, C, D and E interact in a way such that they carry out task F in an optimal manner."* Nonetheless, without generating testable predictions of some form, stating design principles can easily turn into gratuitous theorising. In this thesis, I have made efforts to compare predications derived from the proposed hypotheses with known literature data[‖]. The drawback of only comparing model results to literature data is the risk of, perhaps unconsciously, cherry-picking observations that are consistent with the model. I have, where possible, also proposed experiments that could test the hypotheses stated. However, the proposed experiments are not always straightforward nor will they be absolutely conclusive.

As the experiments I proposed have not yet been performed, some (or all) of the proposed principles can turn out to be wrong. This is not necessarily a bad thing. The design principles proposed here are effectively a logical consequence of what is known about a certain system, combined with reasonable assumptions about their function and the selective pressures acting upon them. As a consequence, when a proposed design principle turns out to be false, this indicates that either our knowledge of the system is incomplete, or our assumptions about the system are incorrect, and we would have learned something in the process. For instance, the proportional relation between ribosomes and growth rate in *E. coli* is found across many different growth conditions. Suppose it turns out that the ppGpp mediated response does *not* robustly lead to this linear relation. This would imply that other mechanisms need to be involved as well, which would also be an interesting finding.

Ultimately, the added value of uncovering design principles lie in their ability to help organise and interpret data about known systems, and provide clues as to

[‖] An example of this is the observation that ppGpp has a strong transient increase upon a nutritional downshift. This data was not used in formulating the model.

how that knowledge can be transferred to other systems. In other words, that we don't need to understand every detail of every (sub-)system to understand the bigger picture. Especially in a discipline as biology, which is characterised by an incredible complexity and diversity, such insights are very useful. Whether such underlying ordering principles are indeed pervasive, or whether they are the exception, is still an open question. The results presented in this thesis provide hope that sometimes the logic underlying complex biological systems is quite simple and general.